Jerk 最小モデルを用いた人の指さし行動予測手法の検討

草 野 祐 樹† 大 村 廉†

近年，様々な生活支援ロボットが研究・開発されてきている．そして，生活支援ロボットは近い将来，日常生活の中に溶け込んでいくことが予想できる．この時，ロボットが人間と『自然』にインタラクションを行えるようになることが，長期にわたるロボットと人間との関係を構築するうえで極めて重要である．本研究では，人間同士のインタラクションにおいて，頻繁に使用される指さし動作に焦点を当てる．そして，3 次元深度画像センサを用い，人間の行動モデルである Jerk 最小モデルに基づき，指さし動作における手先の最終点を予測する手法を提案する．提案する手法を実装して実験を行った．その結果，人間の指さし行動の予測に対して Jerk 最小モデルを適用することの妥当性を確認した．また，予測精度を確認するための実験を行い，位置の平均誤差半径（CEP）は約 180mm であり，動作終了時刻の予測について約 0.5 秒程度の誤差があることを確認した．

A Study on Pointing Gesture Prediction Using Minimum-Jerk Model

Yuki Kusano† and Ren Ohmura†

Recently, many kinds of life support robots have been developed, and the robots are expected to blend with our daily life in near future. Then, naturalness of the human-robot interaction is a key technology for realizing such a future. We focus on human pointing gesture that is one of the most frequent communication methods in our conversation. We propose a method to predict an end of position of a hand by using a 3D depth image sensor and Minimum-Jerk model. We implemented the proposed method and conducted some experiments. The result showed the validity of applying Minimum-Jerk model for the pointing gesture prediction. Also, the position accuracy of the prediction, Circular Error Probability (CEP), was about 180mm, and end time error was about 0.5 seconds.

1. はじめに

近年，様々な生活支援ロボットの研究･開発が多く行われている．近い将来，生活支援ロボットは日常生活の中に溶け込んでいき，人間とロボットが共生する社会になっていくと予想される．そのような社会において，ロボットが人間と『自然』に違和感なくインタラクションを行えるようになることは極めて重要である．例えば，人間同士の会話では，対象を指さしながら会話を進める場合が多くある．ここで，人間とロボットの会話において，人間の指さし行動に対してロボットのレスポンスが遅れてしまうと人間に不自然な印象を与えてしまう．

前述のような対話における指さし行為では，人間は話し相手が指を指す前に大体の方向を無意識に推定している．つまり，他人とやりとりをする時，ある程度先の行動を予測しながら会話などを行う．ロボットと人間の自然なインタラクションを実現するには，この『相手の行動を予測する』ということが必要不可欠である．

本研究では，人間が日常生活で多用しているであろう指さしに焦点を当てる．Sugiyama らは指さしのジェスチャーを予測することが，ロボットと人間との自然なインタラクションを実現するために有効であることを示した[1)]．

Sugiyama らのシステムでは，光学式モーションキャプチャシステムを用い，指の速度，指と体との距離，人間の体の軸と指さしの方向の角度という 3 つのパラメータを用いて，人間の様々な行動の中から指さしだけを認識している．そして，指の速度ベクトルを用いて 0.3 秒後の人間の指の位置の予測を行っている．しかし，Sugiyama らが作成したシステムは指の座標の認識に光学式モーションキャプチャを使用しているため，あらかじめ人間へのマーカーの装着を必要とし，かつ，モーションキャプチャが設置された空間でしか動作しないという問題がある．また，逐次的に予測を行っており，指先が指す軌跡上に複数の物体があった

† 豊橋技術科学大学
Toyohashi University of Technology

場合，ポインティング対象として認識される物体が動作の中で次々に移り変わっていく，という状況が発生する．

そこで本研究では，3 次元深度画像センサ（Microsoft 社の Kinect[2)]）を用いて，初動動作から最終的な指示座標を予測する手法を提案する．現在の 3 次元深度画像センサは，人間と同程度のサイズのロボット（ヒューマノイド）であれば十分に搭載可能な小型のデバイスであり，かつ，人間になんらかの装備を行わずとも人間の姿勢や部位の 3 次元座標を取得することができる．また提案手法では，人間の行動モデルとして Jerk(躍度) 最小モデルを用いた予測を行うことで初動のわずかな動作だけで手の指す最終点を予測する．

以下，第 2 章では本研究の提案手法を提案し，第 3 章でその実装を述べる．そして，第 4 章で本研究の有用性を検討するために行った実験について述べる．最後に，第 5 章で本論文のまとめを述べる．

2. 提案手法

指さし時に手が止まる位置を予測するためには，人間の手の動きがわからなければならない．Flash ら[3)]は平面上の 2 点間を移動する人間の手先の運動を表すモデルとして，Jerk 最小モデルを提案した．Jerk 最小モデルは，完全ではないものの比較的良く人間の運動と一致し，かつ，その軌跡や運動を解析的に解くことができる，という特徴を持つ．本研究では人間の行動モデルとしてこの Jerk 最小モデルを用いる．このモデルは人間の手が移動するとき躍度の積分（式（1））が最小となるというものである．

$$\int_t^T x'''(t)^2 dt \tag{1}$$

そして，このモデルにおいて，人間の手先の動きの開始点，最終点で速度と加速度が 0 であり，開始点を $\boldsymbol{x_0}$，最終点を $\boldsymbol{x_f}$ とすると，ある時刻 τ における手先の座標 $\boldsymbol{x}(\tau)$ は以下のように表すことができる．

$$\boldsymbol{x}(\tau) = \boldsymbol{x_0} + (\boldsymbol{x_f} - \boldsymbol{x_0})(6\tau^5 - 15\tau^4 + 10\tau^3) \tag{2}$$

$$\tau = \frac{t}{T} \tag{3}$$

ここで，t は動作開始時からの時間であり，T は動作開始から終了までにかかる時間である．このとき $\boldsymbol{x}(\tau)$ に対する速度は**図 1** となり，また，加速度は**図 2** のようになる．ここで，加速度の傾きが変化する点 τ_1, τ_2 は，式 (2) を 3 回微分した式

$$\frac{d^3\boldsymbol{x}(\tau)}{d\tau^3} = (\boldsymbol{x_f} - \boldsymbol{x_0})(360\tau^2 - 360\tau + 60) \tag{4}$$

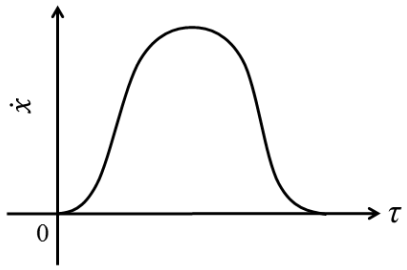

図 1 速度の概形
Fig. 1 Approximate shpe of velocity.

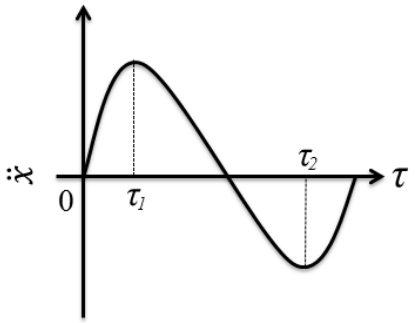

図 2 加速度の概形
Fig. 2 Approximate shpe of accelaration.

から，$\dfrac{d^3\boldsymbol{x}(\tau)}{d\tau^3} = 0$ を満たす τ を求めると

$$\begin{cases} \tau_1 = \dfrac{3-\sqrt{3}}{6} \approx 0.21 \\ \tau_2 = \dfrac{3+\sqrt{3}}{6} \approx 0.79 \end{cases} \tag{5}$$

となる．よって，動作開始から加速度が極大になるまでの時間 τ_1 は動作終了までの時間の約 21%であることがわかる．つまり，指さし行動時に，手先の加速度が増加から減少に変わる (加速度が極大になる瞬間) が分かれば，その指さし行動にかかるおおまかな時間および最終的な手先の座標を求めることができる．終点 $\boldsymbol{x_f}$ を求めるためには式 (2) を

$$\boldsymbol{x_f} = \frac{\boldsymbol{x}(\tau) - \boldsymbol{x_0}}{6\tau^5 - 15\tau^4 + 10\tau^3} + \boldsymbol{x_0} \tag{6}$$

と変形し，その時の手先の座標 $\boldsymbol{x}(\tau_1)$，動作開始時の座標 $\boldsymbol{x_0}$，τ_1(式 (5)) を代入し，

$$\boldsymbol{x_f} = \frac{4\boldsymbol{x}(\tau_1) - (2+\sqrt{3})\boldsymbol{x_0}}{2-\sqrt{3}} \tag{7}$$

を求めることで手先の最終点の座標 $\boldsymbol{x_f}$ を求めることができる．

3. 実　　装

本研究では，手先の軌跡の認識に 3 次元深度画像センサ (Microsoft 社の Kinect) を使用した．そして，

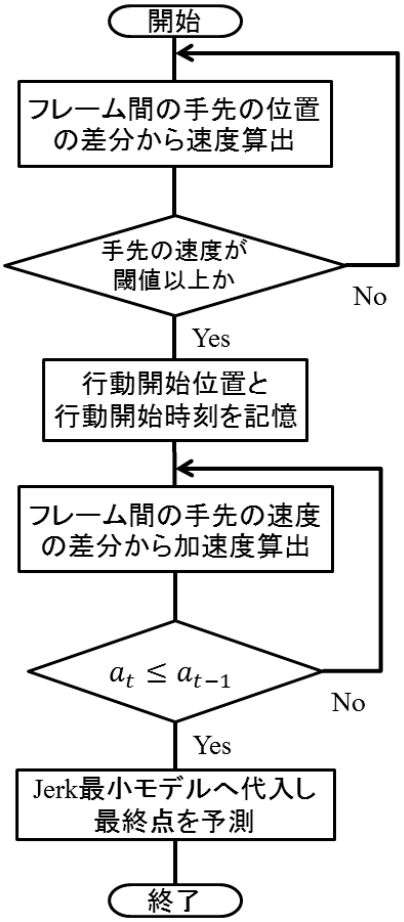

図 3　提案手法を実現するためのアルゴリズム
Fig. 3　The flowchart to achieve proposed method.

Kinect のデバイスコントロールおよび画像処理からのジェスチャ認識を行うために OpenNI[4] を使用して，手先の座標を得た．

提案手法を実現するためのアルゴリズムを**図 3** に示す．指さし行動の開始の認識は以下のように行った．まず，人間の手先の位置をフレームごとに取得し，1 フレーム前の手先の位置との差を求めることにより，フレームごとの速度を求めた．このとき，閾値を決め，その値以上の速度となった時，指さし行動開始時点としてその時刻および手先の座標を記録した．

以降は，手先の位置のフレーム間差分により速度，並びに速度の差分により加速度，加速度の差分により躍度をフレーム毎にそれぞれ求めた．式 (5) の τ_1 に相当する点を検出するために，加速度が極大となる時間を探すが，実際に 0 になることは稀であるため，時刻 t において，指さし行動開始後はじめて

$$a_t \leq a_{t-1} \tag{8}$$

となるとき，$t-1$ を τ_1 の時刻とし，この時の手先の座標を $\boldsymbol{x}(\tau)$ として，式 (7) に代入した．

図 4　システムを用いた指さし動作
Fig. 4　Pointing action using the system.

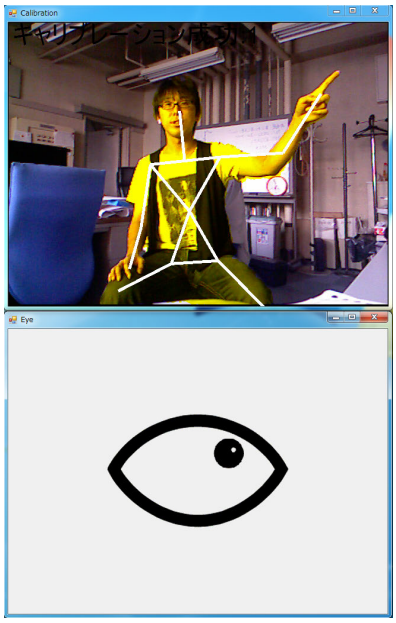

図 5　プログラム実行例
Fig. 5　Sample screen of program.

4. 実　　験

提案手法は現在のところまだ実際のロボットへは適用しておらず，ディスプレイに表示した目を動かすことで動作検証を行っている．実際にプログラムを使って指さしを行っている様子を**図 4** に，作成したプログラムの実行画面例を**図 5** に示す．

提案手法の評価のために 2 種類の実験を行った．まず，Jerk 最小モデルを用いることが妥当かを検証するために実験 1 として，人間が実際に指さし動作を行った時の動作に，Jerk 最小モデルをあてはめることが妥当であるかどうかを確認するための実験を行った．次に，提案手法の有用性を評価するために実験 2 として，提案手法による予測を行ったときにその予測された指の位置と実際に指の止まった位置とどの程度誤差があるか調べるための実験を行った．

なお，実装環境を**表 1** に示す．

表 1　実験環境
Table 1　Experiment environment.

CPU	Intel Core i7-260 3.40GHz
メモリ	4GB
OS	Windows7 64bit
速度の閾値	200mm/s

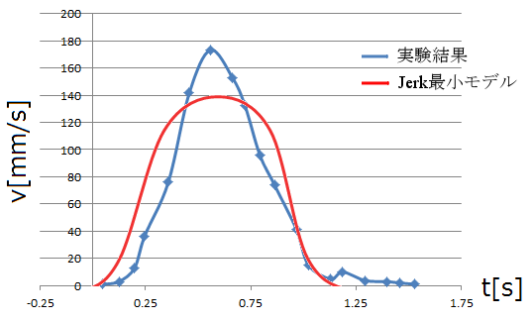

図 6　実験 1 の結果（XYZ 軸を合成した速度）
Fig. 6　Experiment1 result (velocity).

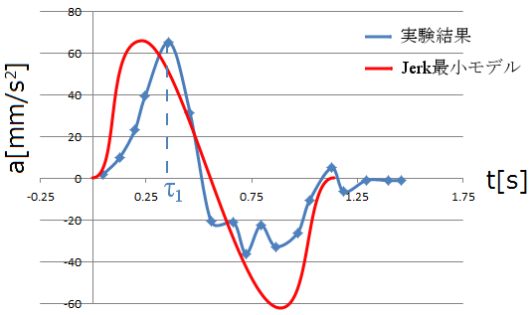

図 7　実験 1 の結果（XYZ 軸を合成した加速度）
Fig. 7　Experiment1 result (accelaration).

4.1　実験 1:Jerk 最小モデル適用の妥当性検証

4.1.1　実 験 方 法

被験者が Kinect から 2.9m 離れ，手を左から右へ移動させ右に 2m のところにあるオブジェクトに指さしを行った．このとき得られた毎フレームの手の位置から速度と加速度のグラフを作成し，それが Jerk 最小モデルに沿うかを検討した．

4.1.2　実 験 結 果

この実験を数回試行したところ，いずれも類似した結果が得られた．また，オブジェクトの位置を変え，左と下に指さしを行ったときも類似した結果が得られた．右に指さしを行った内の一回の結果から得られた速度のグラフと Jerk 最小モデルの速度を図 **6**，加速度のグラフと Jerk 最小モデルの加速度を図 **7** に示す．

4.1.3　考　　察

Jerk 最小モデルに従えば，人間の手の動きの速度は図 1 の形になり，加速度は図 2 の形になる．実験 1 の結果を見ると，速度は図 6 に示すように，少し傾きが大きいが概ね図 1 に近い形をしているといえる．しかし，加速度については図 7 を見ると，グラフが Jerk 最小モデルから少し外れた形となっている．特に，マイナスの値となる後半部分がくずれ，図 2 に示したような形となっていない．しかし，前半部分は比較的良く類似し，かつ，加速度が極大となる時点は Jerk 最小モデル（式（5）の τ_1）に非常に近くなっているといえる．そして，提案手法の予測に用いるのは，τ_1 の時刻とその時刻の位置であることから，提案手法に用いることが可能だと考えた．また，加速度の概形については加速度が 0 からプラスに大きくなり，極大値をとったあとはマイナスの加速度をとるという形になっており図 2 に近いといえる．よって，より詳細な検証が必要であるものの，Jerk 最小モデルを人間の指さしの予測に用いることはある程度の妥当性を有すると考え，以降の実験を進めた．

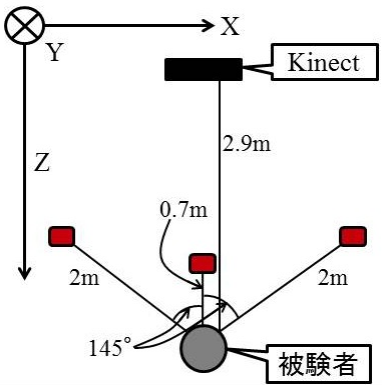

図 8　実験方法
Fig. 8　Experimental method.

4.2　実験 2:提案手法による手首位置予測精度の検証

4.2.1　実 験 方 法

実験 2 では被験者から 2.9m のところに Kinect を設置し，被験者の左右 2m と前方 0.7m の位置にオブジェクトを設置した（図 **8**）．被験者は左右前に置いたオブジェクトに指さしを集中させることによって 3 方向の指さしを各方向 20 回ずつ行った．そして，Jerk 最小モデルによって予測された手の予測停止位置と実際に手の停止した位置との位置の誤差，また，予測した指さしの終了時間と実際の指さしの終了時間との時間の誤差を求めた．

4.2.2　実 験 結 果

実験 2 の実験結果より，XYZ 各軸についての位置誤差の平均と標準偏差と各指さし方向についての平均誤差半径（CEP:Circular Error Probability）☆を求め

☆ 提案手法を用いて予測された手の位置の座標のうち 50%が，実

表 2　位置誤差
Table 2　Position error.

	CEP[mm]	X 軸		Y 軸		Z 軸	
		平均 [mm]	標準偏差	平均 [mm]	標準偏差	平均 [mm]	標準偏差
全体	177.22	-12.83	54.00	28.06	91.80	-122.11	159.04
左	164.79	-24.70	72.89	-7.33	10.01	-158.11	126.32
右	223.29	21.39	15.74	-16.30	31.08	-187.04	139.08
下	139.58	-35.18	37.11	107.81	120.98	-21.19	157.75

表 3　時間誤差
Table 3　Time error.

	T		
	平均 [sec]	標準偏差	誤差率 [%]
全体	0.47	0.47	57.39
左	0.50	0.53	49.17
右	0.53	0.44	72.43
下	0.39	0.42	53.76

た（**表 2**）．その結果，各軸での誤差平均は 200mm 以内であり，XYZ 軸全てを加味した CEP が 177.22mm となった．また，予測された終了時刻についての誤差の平均と標準偏差，誤差率を求めた（**表 3**）．その結果，予測終了時刻が実際の終了時刻より約 0.5 秒遅れているという結果になった．

4.2.3　考　　察

予測した位置と実際に手の停止した位置について，実験 2 から左右の指さしでは Z 軸（奥行き）についての誤差が大きくなるという結果が得られた．これは人間が指さしを行う時，肘や肩を中心として手先が動くため，直線運動にはならず，ベクトルの方向が変化していると考えられる．また，下方向の指さしでは Y 軸（縦方向）の誤差が大きくなるという結果が得られた．その理由としては，人間の腕の長さよりも下に手が動くことはなく，そこで手が急停止するため，この時点で手首の軌跡が Jerk 最小モデルから逸脱してしまうと考えられる．また，図 7 で見られた，時刻 τ_1 と実際の加速度のピークの時刻との差が原因となり，このような誤差が生じたと考えられる．さらに，奥行きの誤差が総じて大きいため，Kinect 自体の手の位置の認識精度による要因も可能性として考えられる．

次に，予測した終了時刻と実際の終了時刻については，予測した終了時刻よりも約 0.5 秒早く指さし動作が完了するという結果となった．この理由として，指さし動作の開始地点は速度の閾値によって検出されるので，閾値が妥当ではないということが考えられる．また，加速度のピーク検出も単純であり，予測が行われる時刻 τ_1 を正確に検出できていない可能性もある．今後の課題となるが，指さし動作開始検出のための速度の閾値の調整，並びに，τ_1 時点検出のための加速度のピーク検出法を工夫することで，より正確な予測が可能になると考えられる．

5. ま　と　め

本論文では人間同士のインタラクションにおいてよく用いられる指さし行動に焦点を当て，ロボットと人間との自然なインタラクションの実現のために，人間にあらかじめ装置などを付与することなく，ロボットが人間の手の指す方向を予測する手法を提案した．提案手法では，3 次元深度画像センサを用い，人間の行動モデルである Jerk 最小モデルに基づき，初動のわずかな動作だけで手の指す最終点を予測することができる．

提案する手法を実装して実験を行い，まず，人の指さし行動の予測に対して Jerk 最小モデルを適用することの妥当性を確認した．そして次に，予測精度を確認するための実験を行った．結果，全体での平均誤差半径（CEP）は約 180mm であり，動作終了時刻の予測については約 0.5 秒程度の誤差があることを確認した．これらは，指さし動作開始検出法および加速度ピーク検出法を工夫することによって改善可能と考えられる．

本研究では，手先の予測までしかできておらず，人間が指さすオブジェクトの特定まではまだ行っていない．今後，動作開始検出および加速度ピーク検出の改善とともに対象物の予測を取り入れ，その検証を行う

際の手の停止位置を中心とした球に納まるときの球の半径である．

予定である．

参 考 文 献

1) O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, N. Hagita: *Natural Deictic Communication with Humanoid Robots*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2007), pp.1441–1448 (2007).
2) http://www.xbox.com/ja-JP/kinect .
3) T. Flash, N. Hogan: *The Coordination of Arm Movements*, J. Neutoscience, vol.5, pp. 1688–1703 (1985).
4) http://www.openni.org/ .